# Higher Order Error of Discrete Fiber Model and Asymptotic Bound on Multistaged PMD Compensation

Yi Li, Avishay Eyal, *Associate Member, IEEE,* and Amnon Yariv, *Life Fellow, IEEE*

***Abstract*—In this paper, we develop a random-matrix formalism that enables analysis of a variety of polarization-mode dispersion (PMD) related problems. In particular, we address the problems of higher order error in a discrete fiber model and limit of multistaged PMD compensation schemes. Our solution to the first problem leads to a simple condition for the validity of the model, which is often overlooked in PMD simulations. For the second issue, we have found an asymptotic bound on the limit of a multistaged PMD compensation scheme. The theory is confirmed by numerical simulations, and future work is suggested.**

***Index Terms*—Jones matrix, optical fiber communications, polarization-mode dispersion (PMD) compensation.**

## I. Introduction

POLARIZATION-MODE dispersion (PMD) has emerged in the last decade as one of the primary obstacles to ultrahigh-speed optical transmission systems and attracted wide interest in optical communications society. It is now well understood that PMD in ordinary optical fibers originates chiefly from random fluctuation of small local birefringence. A mathematically rigorous treatment of the statistical properties of PMD is to directly solve the PMD dynamical equation in the framework of stochastic differential equation (SDE) theory [1]–[5]. The SDE theory is a powerful mathematical tool to deal with the intrinsically stochastic phenomena that naturally arise in, say, statistical physics, quantum optics, noise theory, and financial analysis [6]–[9]. For the particular application to PMD, the SDE method is to give the most realistic description of an ordinary optical fiber and has the advantage of being able to obtain, in principle at least, higher order statistical moments in a systematic manner. Furthermore, it can be readily generalized to account for nonunitary systems as well [10]. All this does not come without a price. The theory of stochastic differential equations is a rather specialized mathematical subject with which many practitioners in the PMD field are not familiar.

If, however, we do not demand as complete a solution as the SDE formalism can offer, then it is possible to circumvent this mathematical difficulty by adopting a "coarse graining" approximation—the discrete random waveplate model (DRW). In this model the fiber is treated as a concatenation of many trunks. Each trunk possesses its own PMD effect and, in a first order approximation, can be treated as an elliptical waveplate. Since the PMD vectors of the subsystems undergo random diffusion over the Poincaré sphere, the optical fiber is visualized in the DRW approximation as a cascade of waveplates undergoing random walk from their initial orientation. Over length and time, the patterns of the waveplates show complete randomness, and it is a common practice to model an optical fiber simply by a cascade of elliptical waveplates whose orientation is randomly distributed over the Poincaré sphere. In fact, this DRW model has been widely employed in PMD simulations as well as in some theoretical investigations [11]–[14], and studies have shown that in the long-length region the DRW model is an excellent description of real optical fibers [4]. A subtle issue of this DRW model worth mentioning is that one needs to use an artificial birefringence that is inversely proportional to the square root of each trunk's length. This is required for the total average PMD of a fiber to remain independent of how many trunks the fiber is divided into. A detailed discussion on this can be found in the appendix of reference [4].

The DRW model is basically a simplified version of the more realistic SDE model, with technical mathematics stripped to minimum. Although there is a general consensus in the field to adopt this simpler, though less realistic, model for a wide range of investigations, the question of its validity and higher order error is not frequently asked. Indeed, as we shall see in Section II, the issue of validity of DRW does arise under certain circumstances. In addition, as illustrated in Section III, the higher order error of the DRW model is intimately related to the analysis of a multistaged PMD compensation protocol.

The rest of this paper is organized as follows. In Section II, a random-matrix formalism for calculating higher order error of the DRW model is developed. In particular, we derive in a closed form the second-order error which suggests a validity condition for the DRW model. We then proceed to apply this formalism to the analysis of a multistaged PMD compensation protocol in Section III. In particular, an asymptotic lower bound on the compensation capacity is found. Section IV contains numerical simulations and Section V summarizes and suggests possible future work.

## II. High-Order Error of DRW

Suppose a single-mode optical fiber is artificially divided into $N$ uncorrelated equal segments. This assumption of uncorrelated segments is valid as long as the length of each segment far

Manuscript received November 30, 1999; revised June 17, 2000. This work was supported by the Office of Naval Research and of the Air Force Office of Scientific Research.

The authors are with the Department of Applied Physics, California Institute of Technology, Pasadena, CA 91125 USA (e-mail: lym@its.caltech.edu).

Publisher Item Identifier S 0733-8724(00)08064-6.

exceeds the fiber diffusion length. If we denote the Jones matrix [15] of the $k$th segment by $\mathbf{T}_k(\omega)$, with $\omega$ denoting the optical frequency, then the Jones matrix of the whole fiber is given by

$$\mathbf{T}_{\rm tot}(\omega) = \mathbf{T}_N(\omega)\cdots\mathbf{T}_2(\omega)\cdot\mathbf{T}_1(\omega). \tag{1}$$

In the following analysis we shall restrict our discussion to the case that no polarization dependent loss/gain (PDL/G) is present. Also we shall assume that a polarization-independent overall phase has been properly separated so that the Jones matrix is in $SU(2)$. That is, the determinant of the Jones matrix is always 1.

The essence of the DRW model is to replace each segment by an elliptical waveplate whose birefringence vector is equal to the PMD vector of that segment. Since a waveplate is strictly a first-order PMD element, it fails to account for the higher order effect in the individual segments. The DRW model can hence be regarded as a first-order approximation for the real fiber. A more realistic model should necessarily consider the higher order PMD effect in each segment, as does the full SDE method which, in fact, takes all orders of effect into account.

To put this on a more concrete footing, we note that the Jones matrix of the, say, $k$th segment can be conveniently expanded in an exponential form [16]

$$\begin{aligned}\mathbf{T}_k(\omega) =& \mathbf{T}_k(\omega_0)\exp\left[\Delta\omega\mathbf{M}_k^{(1)}\right]\cdot\exp\left[\tfrac{1}{2!}\,\Delta\omega^2\mathbf{M}_k^{(2)}\right]\\ &\times\exp\left[\tfrac{1}{3!}\,\Delta\omega^3\mathbf{M}_k^{(3)}\right]\cdots\end{aligned} \tag{2}$$

where $\Delta\omega := \omega-\omega_0$ is the frequency detuning from the carrier frequency $\omega_0$. A brief discussion of this exponential expansion and several useful identities can be found in the Appendix. For now it is sufficient to note that matrix $\mathbf{M}_k^{(n)}$ is a constant $2\times 2$ matrix that, roughly speaking, describes the $n$th order PMD effect of the $k$th segment. In a DRW approximation, all product terms higher than the first order in (2) are neglected, and $\mathbf{T}_k(\omega)$ is replaced by the Jones matrix of the corresponding waveplate

$$\mathbf{P}_k(\omega) = \mathbf{T}_k(\omega_0)\exp\left[\Delta\omega\mathbf{M}_k^{(1)}\right]. \tag{3}$$

The Jones matrix of the whole fiber is thus approximated by that of the DRW

$$\mathbf{P}_{\rm tot}(\omega) = \mathbf{P}_N(\omega)\cdots\mathbf{P}_2(\omega)\cdot\mathbf{P}_1(\omega). \tag{4}$$

Equations (1)–(4) are our starting point to calculate the higher order error of the DRW model. Before we start, some comments are in order here. We will study only the cases that the DRW model is approximately valid. That is, each segment can, to a good degree, be well approximated by an elliptical waveplate. Higher order effect in individual segments can, therefore, be treated as small perturbations. In fact, this condition is already implicitly inferred by the DRW model itself—for the latter to work *at all*, the individual segments have to behave quite similar to elliptical waveplates. This is by no means saying that we are only to calculate small corrections to the DRW model. The small perturbations in each of the $N$ segments may add up to cause significant discrepancy of the DRW model from the real fiber. It is this potentially significant *total* error that we are after.

In the following, we shall demonstrate how to calculate the second-order error. In many situations this is sufficient for assessing the validity of the DRW model. Furthermore, for sufficiently large $N$—such as in fiber simulations—the dominant error is in the second order. It should be noted that, however, the formalism developed below is completely general and can be carried to higher orders.

To second order, (2) can be written as

$$\begin{aligned}\mathbf{T}_k(\omega) \simeq& \mathbf{P}_k(\omega)\exp\left[\tfrac{1}{2}\,\Delta\omega^2\mathbf{M}_k^{(2)}\right]\\ \simeq& \mathbf{P}_k\left(\mathbf{1}+\tfrac{1}{2}\,\Delta\omega^2\mathbf{M}_k^{(2)}\right).\end{aligned} \tag{5}$$

Inserting (5) into (1) and keeping only terms up to $\mathcal{O}(\Delta\omega^2\mathbf{M}_k^{(2)})$ we obtain

$$\begin{aligned}\mathbf{T}_{\rm tot}(\omega) \simeq& \mathbf{P}_N\left(\mathbf{1}+\frac{\Delta\omega^2}{2}\,\mathbf{M}_N^{(2)}\right)\cdots\mathbf{P}_1\left(\mathbf{1}+\frac{\Delta\omega^2}{2}\,\mathbf{M}_1^{(2)}\right)\\ \simeq& \mathbf{P}_{\rm tot}(\omega)+\tfrac{1}{2}\,\Delta\omega^2\sum_{k=1}^{N}\mathbf{P}_N\cdots\\ &\times\mathbf{P}_k\mathbf{M}_k^{(2)}\mathbf{P}_{k-1}\cdots\mathbf{P}_2\mathbf{P}_1\\ =& \mathbf{P}_{\rm tot}(\omega)+\tfrac{1}{2}\,\Delta\omega^2\sum_{k=1}^{N}\mathbf{G}_k\\ \mathbf{G}_k :=& \mathbf{P}_N\cdots\mathbf{P}_k\mathbf{M}_k^{(2)}\mathbf{P}_{k-1}\cdots\mathbf{P}_2\mathbf{P}_1.\end{aligned} \tag{6}$$

Here, $\mathbf{G}_k$'s are introduced to simplify the notation. We next define the error matrix $\mathbf{D}(\omega) := \mathbf{T}_{\rm tot}(\omega)-\mathbf{P}_{\rm tot}(\omega)$. By (6), we have:

$$\mathbf{D}(\omega) \simeq \tfrac{1}{2}\,\Delta\omega^2\sum_{k=1}^{N}\mathbf{G}_k. \tag{7}$$

On the other hand, since we can write the unitary matrices $\mathbf{T}_{\rm tot}$ and $\mathbf{P}_{\rm tot}$ in the form

$$\mathbf{T}_{\rm tot},\,\mathbf{P}_{\rm tot} = \begin{bmatrix} a_{T,P} & b_{T,P}\\ -b_{T,P}^* & a_{T,P}^*\end{bmatrix} \tag{8}$$

where $|a_T|^2+|b_T|^2 = 1 = |a_P|^2+|b_P|^2$, the error matrix assumes the following form:

$$\mathbf{D}(\omega) = \begin{bmatrix}\Delta a & \Delta b\\ -\Delta b^* & \Delta a^*\end{bmatrix} \tag{9}$$

where $\Delta a = a_T-a_P$, $\Delta b = b_T-b_P$. The DRW error function, defined by

$$f(\omega) := |\Delta a(\omega)|^2+|\Delta b(\omega)|^2 \tag{10}$$

is a measure of the error of DRW and can be expressed in terms of $\mathbf{G}_k$

$$f(\omega) = \det|\mathbf{D}(\omega)| \simeq \tfrac{1}{4}\,\Delta\omega^4\det\left|\sum_{k=1}^{N}\mathbf{G}_k\right|. \tag{11}$$

To relate $f$ to the usual PMD parameters of the fiber, let us review a few PMD identities here for later convenience. The PMD effect in an optical fiber is described by the PMD vector, $\mathbf{\Omega}$. It is, in general, frequency dependent. Its magnitude at some

fixed carrier frequency is called the first-order PMD at that frequency and is denoted by $\Omega$. The second-order PMD, denoted by $\Omega_\omega$, is defined by $\Omega_\omega := |\partial\mathbf{\Omega}/\partial\omega|$. The mean-square first-order and second-order PMD satisfy the following well-known identity [1]:

$$\langle\Omega_\omega^2\rangle = \tfrac{1}{3}\langle\Omega^2\rangle^2. \tag{12}$$

Also, if the fiber is divided into $N$ equal segments, and we denote the first-order PMD of one segment and the whole fiber by $\Omega$ and $\Omega_{\rm tot}$, respectively, then they are related by

$$\langle\Omega_{\rm tot}^2\rangle = N\langle\Omega^2\rangle. \tag{13}$$

Let us now return to our calculation of $f$ in (11). In the Appendix, it is shown that $\mathbf{M}_k^{(2)}$ is proportional to a unitary matrix, and the proportionality factor, $\lambda_k$, is one half the magnitude of the second-order PMD of the $k$th segment: $\lambda_k = \Omega_{\omega,\,k}/2$. It follows from the definition of $\mathbf{G}_k$ in (6) that $\mathbf{G}_k$ can be written in the form of $\mathbf{G}_k = \lambda_k\mathbf{U}_k$ for some unitary $\mathbf{U}_k$. Adopting this new notation, we have

$$f(\omega) \simeq \tfrac{1}{4}\,\Delta\omega^4 \det\left|\sum_{k=1}^{N}\lambda_k\mathbf{U}_k\right|. \tag{14}$$

To carry out the calculation further, it is beneficial to express $\mathbf{U}_k$ explicitly in the following form:

$$\mathbf{U}_k = \begin{bmatrix} \cos\theta_k e^{i\varphi_k} & \sin\theta_k e^{i\phi_k} \\ -\sin\theta_k e^{-i\phi_k} & \cos\theta_k e^{-i\varphi_k} \end{bmatrix} \tag{15}$$

where $\theta_k \in [0,\,\pi/2]$, $\varphi_k \in [0,\,2\pi)$ and $\phi_k \in [0,\,2\pi)$. Also, needed are the *interweaved matrices*:

$$\mathbf{U}_{jk} := \begin{bmatrix} \cos\theta_j e^{i\varphi_j} & \sin\theta_k e^{i\phi_k} \\ -\sin\theta_j e^{-i\phi_j} & \cos\theta_k e^{-i\varphi_k} \end{bmatrix}. \tag{16}$$

Intuitively, the first column of $\mathbf{U}_{jk}$ is the first column of $\mathbf{U}_j$, and the second column of $\mathbf{U}_{jk}$ is the second column of $\mathbf{U}_k$. Then by the theory of linear algebra

$$\begin{aligned}\det\left|\sum_{k=1}^{N}\lambda_k\mathbf{U}_k\right| &= \sum_{j=1}^{N}\sum_{k=1}^{N}\lambda_j\lambda_k\det|\mathbf{U}_{jk}| \\ &= \sum_{k=1}^{N}\lambda_k^2\det|\mathbf{U}_{kk}| + \sum_{1\le j<k\le N}\lambda_j\lambda_k \\ &\quad\times(\det|\mathbf{U}_{jk}| + \det|\mathbf{U}_{kj}|).\end{aligned} \tag{17}$$

Noting that $\det|\mathbf{U}_{kk}| = \det|\mathbf{U}_k| = 1$, the first summation on the right-hand side (RHS) of (17) is simply $\sum_{k=1}^{N}\lambda_k^2 = \sum_{k=1}^{N}\Omega_{\omega,\,k}^2/4$. For large $N$, we can replace $\Omega_{\omega,\,k}^2$ by its average value $\langle\Omega_{\omega,\,k}^2\rangle := \langle\Omega_\omega^2\rangle$ which is independent of the segment index $k$. By virtue of (12) and (13), it follows:

$$\sum_{k=1}^{N}\lambda_k^2\det|\mathbf{U}_{kk}| = \frac{\langle\Omega_{\rm tot}^2\rangle^2}{12N}. \tag{18}$$

To calculate the second summation in (17), we note that

$$\begin{aligned}u_{jk} &:= \det|\mathbf{U}_{jk}| + \det|\mathbf{U}_{kj}| \\ &= \cos\theta_j\cos\theta_k e^{i(\varphi_j-\varphi_k)} + \sin\theta_j\sin\theta_k e^{i(\phi_k-\phi_j)} + \text{c.c.} \\ &= 2\cos\theta_j\cos\theta_k\cos(\varphi_j-\varphi_k) \\ &\quad + 2\sin\theta_j\sin\theta_k\cos(\phi_k-\phi_j).\end{aligned} \tag{19}$$

As the individual segments are assumed to be uncorrelated, $\theta_{j,\,k}$, $\varphi_j-\varphi_k$ and $\phi_j-\phi_k$ can be regarded as independent uniform random variables over intervals $[0,\,\pi/2]$, $[0,\,2\pi)$ and $[0,\,2\pi)$, respectively. Therefore $u_{jk}$ so defined is also a random variable with mean

$$\langle u_{jk}\rangle = 0 \tag{20}$$

and variance

$$\begin{aligned}\mathrm{Var}(u_{jk}) &= \langle u_{jk}^2\rangle - \langle u_{jk}\rangle^2 \\ &= 4\langle\cos^2\theta_j\rangle\langle\cos^2\theta_k\rangle\langle\cos^2(\varphi_j-\varphi_k)\rangle \\ &\quad + 4\langle\sin^2\theta_j\rangle\langle\sin^2\theta_k\rangle\langle\cos^2(\phi_j-\phi_k)\rangle = 1.\end{aligned} \tag{21}$$

It then follows that the summand $\lambda_j\lambda_k(\det|\mathbf{U}_{jk}| + \det|\mathbf{U}_{kj}|)$ in (17) can be regarded as a random variable with mean zero and variance $\lambda_j^2\lambda_k^2$. In addition, they are independent of each other. By the central limit theorem of probability, the second summation in (17) is approximately a Gaussian random variable with mean zero and standard deviation

$$\begin{aligned}\sigma &= \left[\sum_{j<k}\lambda_j^2\lambda_k^2\right]^{1/2} \simeq \left[\tfrac{1}{2}\,N^2\langle\lambda_j^2\rangle\langle\lambda_k^2\rangle\right]^{1/2} \\ &= \frac{N}{4\sqrt{2}}\langle\Omega_\omega^2\rangle = \frac{1}{12\sqrt{2}N}\langle\Omega_{\rm tot}^2\rangle^2\end{aligned} \tag{22}$$

in the case of sufficiently large $N$. It is evident from (18) and (22) that the two sums in (17) have comparable contributions to the second-order correction. However, it is also clear that the second sum is a randomly fluctuation. In a time average (or an ensemble average) only the first sum survives. According to (14), (17), and (18), we obtain the averaged error function:

$$\langle f(\omega)\rangle \simeq \frac{\Delta\omega^4\langle\Omega_{\rm tot}^2\rangle^2}{48N} \tag{23}$$

and its frequency average over the entire transmission bandwidth

$$\langle f\rangle \simeq \frac{\Delta\omega_0^4\langle\Omega_{\rm tot}^2\rangle^2}{240N}. \tag{24}$$

In (24) $\Delta\omega_0$ is the *half* frequency bandwidth, and we have assumed that the carrier frequency $\omega_0$ lies at the center of the band.

Equation (24) is one of our key results in this paper. Recalling the definition of $f$ in (10) we reach the following important conclusion: *the root-mean-square (rms) error of the DRW model decays as $1/\sqrt{N}$, with a linear coefficient depending solely on a*

*dimensionless system-dependent constant* $\kappa := 2\Delta\omega_0\langle\Omega_{\rm tot}^2\rangle^{1/2}$ *it or, mathematically,*

$$\langle|\Delta a|^2+|\Delta b|^2\rangle^{1/2} \simeq \frac{\kappa^2}{16\sqrt{15N}}. \tag{25}$$

The question of how accurately the DRW model reproduces a real fiber can now be answered. For a faithful approximation, the RHS of (25) must be small:

$$\frac{\kappa^2}{16\sqrt{15N}} \ll 1. \tag{26}$$

Under this condition, the DRW model is a faithful representation for a realistic fiber. We note in passing that in PMD simulations using DRW, it is always a good practice to check if (26) is satisfied. We remark that the validity of the DRW model can almost always be ensured in the sense of (26) by suitably choosing a large $N$ if the modeled fiber is long enough.

## III. Asymptotic Bound on Multistaged PMD Compensation

Multistaged PMD compensation (MPC) schemes have recently attracted attention as a potential candidate for effective PMD suppression in future ultrafast long-distance optical transmission systems [17], [18]. The main merit of MPC exists in its great capability of mitigating severe PMD effect which may cripple the usual first-order and second-order compensation methods. In our model of MPC, the compensator is composed of a large number of adjustable birefringent elements. Hereafter, we will refer to them simply as waveplates, although in reality they may take a variety of physical forms that bear no resemblance to ordinary waveplates. The axes and the magnitudes of birefringence of these waveplates can both be adaptively controlled by feedback mechanisms in order that the PMD of the combined system—the fiber and the compensator together—is minimized.

As in the previous section, let $\mathbf{T}_{\rm tot}(\omega)$ be the Jones matrix of the optical fiber link. Ideally we would like to build a compensator that eliminates the PMD completely. The Jones matrix of such an ideal compensator is necessarily $\mathbf{T}_{\rm tot}^{-1}(\omega)$, up to a constant unitary rotation which can be eliminated by a proper choice of basis. In our MPC model, however, the compensator contains only a finite number of waveplates, so its Jones matrix is a function of a finite number of parameters, such as the orientational angles of the waveplates on the Poincaré sphere. By function theory, a function of finite number of parameters cannot reproduce an *arbitrary* continuous function [in our case, $\mathbf{T}_{\rm tot}^{-1}(\omega)$]. Therefore, no perfect PMD compensator that employs only a finite number of waveplates exists.

Since no perfect compensation exists, any compensation scheme will inevitably leave some residue PMD. The *performance figure* of a PMD compensator is defined as

$$Q := \left[\frac{\langle\Omega_{\rm tot}^2\rangle}{\langle\Omega_{\rm res}^2\rangle}\right]^{1/2} \tag{27}$$

with $\langle\Omega_{\rm tot}^2\rangle$ and $\langle\Omega_{\rm res}^2\rangle$ stand, respectively, for the mean-square original and residue PMD. The higher the $Q$ value is, the better the performance of the PMD compensator. In realistic MPC devices, PMD compensation is achieved by real-time optimization control of the waveplates in the compensator to maximize a pre-selected merit function. Even without considering experimental errors, different choices of the merit function and control procedure will in general yield different $Q$ values for the same MPC model. We say that the performance figure is *algorithm-dependent*. A quantity that is closely related to $Q$ and depends only on the intrinsic structure of the MPC model is the *compensation capacity* $Q_c$, which is defined by the upper limit of $Q$ over all possible control procedures and merit function choices. The compensation capacity $Q_c$ is analogous to the channel capacity in information theory [19]. In information theory, the transmission rate over a noisy channel depends on the specific choice of coding and decoding procedure, but the maximum achievable rate is set by the channel capacity, which depends only on the properties of the channel itself.

It would be of great interest to explicitly calculate $Q_c$ for a given MPC protocol (such as ours discussed above). Unfortunately we have not found an analytical result by the time of this writing. Nevertheless, we have calculated a lower bound on $Q_c$ and, in the limit of large $N$, it may serve as a good estimate.

Suppose the compensator contains $N$ adjustable waveplates. Accordingly, we divide the fiber into $N$ segments and write its Jones matrix as $\mathbf{T}_{\rm tot} = \mathbf{T}_N\cdots\mathbf{T}_2\cdot\mathbf{T}_1$ and the Jones matrix of its corresponding DRW model as $\mathbf{P}_{\rm tot} = \mathbf{P}_N\cdots\mathbf{P}_2\cdot\mathbf{P}_1$, as in Section II. Again we make the assumption that the length of each segment is much larger than the fiber's diffusion length. Since in our model we are free to adjust the waveplates' orientation on the *entire* Poincaré sphere, as well as their magnitude, the compensator can, in principle, be configured to generate $\mathbf{P}_{\rm tot}^{-1} = \mathbf{P}_{\rm tot}^{\dagger} = \mathbf{P}_1^{\dagger}\cdot\mathbf{P}_2^{\dagger}\cdots\mathbf{P}_N^{\dagger}$ to arbitrary precision. The performance figure $Q_0$ associated with this particular realization is obviously a lower bound for $Q_c$: $Q_c \geq Q_0$. On the other hand, since $\mathbf{P}_{\rm tot}^{-1} \to \mathbf{T}_{\rm tot}^{-1}$ for large $N$, this particular realization is close to any optimal realization in the asymptotic limit of large $N$,[1] and $Q_0$ may be used in place of $Q_c$ for rough estimates in the design of such MPC devices, as we shall show by the end of this section. In the following, we will explicitly calculate $Q_0$ in a second-order treatment.

Taking the Jones matrix of the compensator to be $\mathbf{P}_{\rm tot}^{-1}$, we can write the Jones matrix of the combined system—the fiber and the compensator together—as

$$\begin{aligned}\mathbf{R}(\omega) &= \mathbf{P}_{\rm tot}^{\dagger}(\omega)\cdot\mathbf{T}_{\rm tot}(\omega)\\ &\simeq \mathbf{1} + \frac{\Delta\omega^2}{2}\sum_{k=1}^{N}\mathbf{P}_1^{\dagger}\cdots\mathbf{P}_{k-1}^{\dagger}\mathbf{M}_k^{(2)}\mathbf{P}_{k-1}\cdots\mathbf{P}_1\\ &= \mathbf{1} + \tfrac{1}{2}\,\Delta\omega^2\sum_{k=1}^{N}\mathbf{H}_k,\end{aligned}$$

where

$$\mathbf{H}_k := \mathbf{P}_1^{\dagger}\cdots\mathbf{P}_{k-1}^{\dagger}\mathbf{M}_k^{(2)}\mathbf{P}_{k-1}\cdots\mathbf{P}_1. \tag{28}$$

[1]The validity of this statement is subject to the choice of MPC models. In our protocol, the waveplates can be freely rotated over the entire Poincaré sphere. The statement does not apply to models that utilize linear waveplates. We shall restrict our discussion to the former.

Here, we have invoked (6) from Section II and introduced $\mathbf{H}_k$ to simplify the notation. The residue PMD after compensation, $\Omega_{\rm res}$, then satisfies

$$\frac{\Omega_{\rm res}^2(\omega)}{4} = \det|\mathbf{R}'(\omega)| = \det\left|\Delta\omega\sum_{k=1}^{N}\mathbf{H}_k + \frac{\Delta\omega^2}{2}\sum_{k=1}^{N}\mathbf{H}_k'\right|. \quad (29)$$

A simple derivation of (29) is given at the end of the Appendix.

The determinant that appears in (29) can be expanded into a sum of determinants similar to that in (17), with the only difference that there are now three distinct groups of terms: those which correlate among $\mathbf{H}_k$s, those which correlate among $\mathbf{H}_k'$s, and those which correlate *both* $\mathbf{H}_j$s and $\mathbf{H}_k$s. By definition of $\mathbf{H}_k$ in (28) and the basic assumption of the DRW model, it is rational to treat $\mathbf{H}_k$ as completely uncorrelated to $\mathbf{H}_j'$. By an argument similar to what leads to (22) in the previous section, the contribution of the third group can be regarded as a random fluctuation. Writing this out explicitly, we have

$$\frac{\Omega_{\rm res}^2(\omega)}{4} = \det\left|\Delta\omega\sum_{k=1}^{N}\mathbf{H}_k\right| + \det\left|\frac{\Delta\omega^2}{2}\sum_{k=1}^{N}\mathbf{H}_k'\right| + \text{r.f.t.} \quad (30)$$

where "r.f.t." represents "randomly fluctuating term"—it is a term that vanishes under time or ensemble averaging.

Next, we note that the first term in (30) is exactly similar to what we have encountered in (11) in Section II. A completely parallel calculation yields

$$\det\left|\Delta\omega\sum_{k=1}^{N}\mathbf{H}_k\right| = \frac{\Delta\omega^2\langle\Omega_{\rm tot}^2\rangle^2}{12N} + \text{r.f.t.} \quad (31)$$

We now focus on the second term in (30). First, we define a new matrix

$$\mathbf{H}_{km} := \mathbf{P}_1^\dagger\mathbf{P}_2^\dagger\cdots\mathbf{P}_{k-1}^\dagger\mathbf{M}_k^{(2)}\mathbf{P}_{k-1}\cdots\left(\frac{\partial\mathbf{P}_m}{\partial\omega}\right)\times\mathbf{P}_{m-1}\cdots\mathbf{P}_2\mathbf{P}_1, \qquad m<k. \quad (32)$$

It can be shown (see the Appendix) that $\mathbf{M}_k^{(2)\dagger} = -\mathbf{M}_k^{(2)}$. Therefore

$$\mathbf{H}_{km}^\dagger = -\mathbf{P}_1^\dagger\mathbf{P}_2^\dagger\cdots\left(\frac{\partial\mathbf{P}_m^\dagger}{\partial\omega}\right)\mathbf{P}_{m+1}^\dagger\cdots\mathbf{P}_{k-1}^\dagger\mathbf{M}_k^{(2)}\times\mathbf{P}_{k-1}\cdots\mathbf{P}_2\mathbf{P}_1. \quad (33)$$

It follows from (32) and (33) that:

$$\mathbf{H}_k' = \sum_{m=1}^{k-1}(\mathbf{H}_{km} - \mathbf{H}_{km}^\dagger). \quad (34)$$

As shown in the Appendix, both $\mathbf{M}_m^{(1)}$ and $\mathbf{M}_k^{(2)}$ are unitary matrices scaled by positive factors $\alpha_m$ and $\lambda_k$, with $\alpha_m = \Omega_m/2$ and $\lambda_k = (\Omega_{\omega,k}/2)$. By the definition of $\mathbf{H}_{km}$ and noting that $\partial\mathbf{P}_m/\partial\omega = \mathbf{P}_m\mathbf{M}_m^{(1)}$, we can write $\mathbf{H}_{km} = \alpha_m\lambda_k\mathbf{U}_{km}$ for some unitary $\mathbf{U}_{km}$. Then second term in (30) is reduced to

$$\begin{aligned}\det\left|\sum_{k=1}^{N}\mathbf{H}_k'\right| &= \det\left|\sum_{k=1}^{N}\sum_{m=1}^{k-1}(\mathbf{H}_{km}-\mathbf{H}_{km}^\dagger)\right| \\ &= \sum_{k=1}^{N}\sum_{m=1}^{k-1}\det\left|(\mathbf{H}_{km}-\mathbf{H}_{km}^\dagger)\right| + \text{r.f.t.} \\ &= \sum_{k=1}^{N}\sum_{m=1}^{k-1}\frac{\Omega_m^2\Omega_{\omega,k}^2}{16}\det\left|(\mathbf{U}_{km}-\mathbf{U}_{km}^\dagger)\right| + \text{r.f.t.}\end{aligned} \quad (35)$$

In the second step above, we have used the same technique as applied in Section II, by treating $(\mathbf{H}_{km} - \mathbf{H}_{km}^\dagger)$'s as uncorrelated matrices, to decompose the determinant of the matrix sum into a positive term and a randomly fluctuating term.

Next, we write $\mathbf{U}_{km}$ explicitly as

$$\mathbf{U}_{km} = \begin{bmatrix}\cos\theta_{km}e^{i\phi_{km}} & \sin\theta_{km}e^{i\varphi_{km}} \\ -\sin\theta_{km}e^{-i\varphi_{km}} & \cos\theta_{km}e^{-i\phi_{km}}\end{bmatrix} \quad (36)$$

and

$$\det\left|(\mathbf{U}_{km}-\mathbf{U}_{km}^\dagger)\right| = 4\cos^2\theta_{km}\sin^2\phi_{km} + 4\sin^2\theta_{km}. \quad (37)$$

By treating $\theta_{km}$ and $\phi_{km}$ as uniform random variables, we can regard $\det|(\mathbf{U}_{km} - \mathbf{U}_{km}^\dagger)|$ as a random variable with mean $\langle 4\cos^2\theta_{km}\sin^2\phi_{km} + 4\sin^2\theta_{km}\rangle = 3$. Hence, for large $N$ the double sum in the last line of (35) can be approximated by

$$\begin{aligned}&\sum_{k=1}^{N}\sum_{m=1}^{k-1}\frac{\Omega_m^2\Omega_{\omega,k}^2}{16}\det\left|(\mathbf{U}_{km}-\mathbf{U}_{km}^\dagger)\right| \\ &\quad\simeq \sum_{k=1}^{N}\sum_{m=1}^{k-1}\left\langle\frac{\Omega_m^2\Omega_{\omega,k}^2}{16}\det\left|(\mathbf{U}_{km}-\mathbf{U}_{km}^\dagger)\right|\right\rangle \\ &\quad= \frac{1}{32}N(N-1)\langle\Omega^2\rangle\langle\Omega_\omega^2\rangle\cdot 3 \simeq \frac{\langle\Omega_{\rm tot}^2\rangle^3}{32N}.\end{aligned} \quad (38)$$

Here, we have used the fact that for $k \neq m$, $\langle\Omega_m^2\Omega_{\omega,k}^2\rangle = \langle\Omega_m^2\rangle\langle\Omega_{\omega,k}^2\rangle$, and the familiar relations given by (12) and (13). From (31), (35), and (38), (30) reduces to

$$\Omega_{\rm res}^2(\omega) = \frac{\Delta\omega^2\langle\Omega_{\rm tot}^2\rangle^2}{3N} + \frac{\Delta\omega^4\langle\Omega_{\rm tot}^2\rangle^3}{32N} + \text{r.f.t.} \quad (39)$$

After taking the time (or ensemble) average as well as the frequency average over the transmission bandwidth on both sides of (39) and remembering the definition of the performance figure (27), we finally obtain

$$Q_0 = \left[\left(\frac{\kappa^2}{36} + \frac{\kappa^4}{640}\right)\frac{1}{N}\right]^{-(1/2)} \quad (40)$$

where $\kappa := 2\Delta\omega_0\langle\Omega_{\rm tot}^2\rangle^{1/2}$ as previously defined.

After obtaining a lower bound for $Q_c$, it is natural to look for an upper bound as well. Unfortunately, we have not found one yet and it remains a problem to be solved.

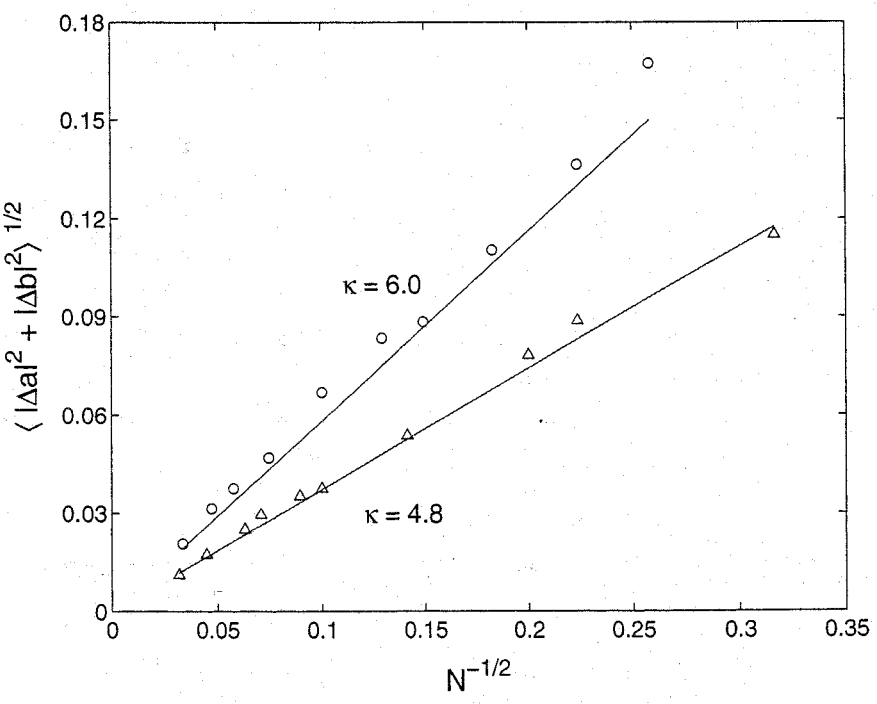


Fig. 1. Higher order error of the DRW model for a single-mode optical fiber. Numerical calculations (triangles and circles) versus theoretical prediction (25) (solid lines).

A more convenient measure of the residue PMD is the PMD reduction ratio $\rho$, defined as the inverse of the performance figure [refer to (27)], and its *minimum* attainable value is, $\rho_c = Q_c^{-1}$, has an upper bound

$$\rho_0 := \frac{1}{Q_0} = \left(\frac{\kappa^2}{36} + \frac{\kappa^4}{640}\right)^{1/2} \frac{1}{\sqrt{N}}. \tag{41}$$

Equation (41) is useful for assessing some design issues. For example, in order to achieve a prespecified PMD reduction ratio $\rho_s$, the minimum required number of waveplates can be estimated, by setting $\rho_s \sim \rho_0$, to be

$$N_s \sim \left(\frac{\kappa^2}{36} + \frac{\kappa^4}{640}\right) \frac{1}{\rho_s^2}. \tag{42}$$

The above estimate on $N_s$ is expressed solely as a function of $\rho_s$ and $\kappa$. For transmission links with moderate PMD effect ($\kappa \sim 1$), $N_s$ is bounded above quadratically in $\kappa$. In situations subject to severe PMD effect ($\kappa \gg 1$), it is only bounded quadruply. The increasing difficulty in equalization as PMD is getting larger can be appreciated from this observation.

## IV. Numerical Results

Numerical simulations are performed to check the validity of various assumptions made in our theoretical calculation in the previous two sections. In particular, we compare the theoretical predictions of (25) and (41) with numerical calculations. For this purpose, numerical results calculated from the stochastic model [1], [3], [4] are used as standard values with which the theoretical predictions are compared. Here we shall follow Wai and Menyuk [4] closely. The numerical model assumes constant strength of local linear birefringence and randomly varying orientation of local birefringence axes over length that can be represented by an Ornstein–Uhlenbeck process [4]

$$\left\langle \frac{d\theta}{dz} \right\rangle = 0 \qquad \left\langle \frac{d\theta}{dz}\,\frac{d\theta}{dz'} \right\rangle = \frac{2}{h}\,\delta(z - z') \tag{43}$$

where $\theta$ is the azimuthal angle of the local birefringence axis in the linear plane of the Poincaré sphere and $h$ is the autocorrelation length of the fiber. The assumption of a constant local birefringence strength is not restrictive, since a more elaborate model, which allows for random variation of the strength yields essentially the same results [4].

The simulations are performed on two sets of fibers with the following different transmission characteristics:

| | Set 1 | Set 2 |
|---|---|---|
| $\Delta n$ | $5.0 \times 10^{-6}$ | $1.0 \times 10^{-7}$ |
| $h$ | 20 m | 50 m |
| $L$ | 900 km | 1000 km |
| $2\Delta\omega_0/2\pi$ | 9.55 GHz | 229 GHz |
| $\kappa$ | 6.0 | 4.8. |

Here $\Delta n$ is the strength of the local birefringence, and $L$ the total length of the fiber. Note that intermediate $\kappa$ values are chosen so that the quadratic and quadruple terms in (41) have comparable magnitudes. In both models the step length of calculation is 1 m, far less than the autocorrelation length $h$ to ensure numerical accuracy.

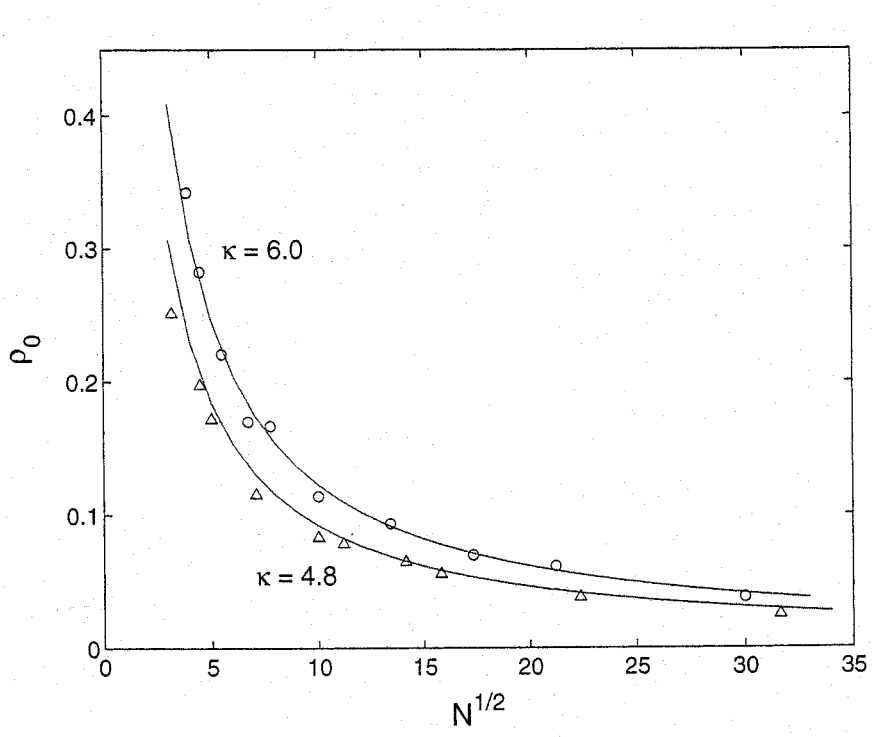


Fig. 2. Upper bound $\rho_0$ on the average PMD reduction ratio in our MPC protocol. Numerical calculations (circles and triangles) versus theoretical prediction (41) (solid lines).

In each realization of the fiber, the Jones matrix of the corresponding DRW model, with $N$ ranging from 10 to 1000, is compared to the "actual" Jones matrix given by the stochastic model. The errors are plotted in Fig. 1, with circles for set-1 and triangles for set-2. Theoretical prediction of the second-order error given by (25) is plotted as solid lines. The agreement verifies the assumption that the second-order error is dominant. It also suggests that the random-matrix approach in our calculation is valid.

Also calculated are numerical values of $\rho_0$. In each fiber realization, $\mathbf{P}_{\rm tot}^{-1}$ is calculated and so is the corresponding residue PMD $\langle\Omega_{\rm res}^2\rangle$, from which numerical values of $\rho_0$ are obtained. These are plotted in Fig. 2, with circles for set-1 and triangles for set-2. Theoretical plots of (41) are the solid curves. In both cases of Figs. 1 and 2, the numerical data are averaged over 200 independent realizations. The agreement in both cases is gratifying.

## V. Conclusion

In this paper, we develop a random-matrix formalism that can be used to study a variety of PMD-related problems. In Section II we calculate the second-order error of the DRW model that has direct implication for PMD simulations using DRW. In Section III the same formalism is used to calculate a lower bound on the compensation capacity of an MPC protocol. In both cases, the results depend only on system parameters $N$ and $\kappa$. The theoretical results are verified by numerical simulations.

Regarding the compensation capacity of an MPC protocol, there are many open problems that awaits resolution. We will list a few:

1) finding an *upper* bound for $Q_c$;
2) or, better yet, calculating $Q_c$ explicitly;
3) a parallel analysis for other MPC models, such as those implementing only linear waveplates.

The answers to these (and other) questions are likely to enhance our theoretical understanding of multistaged PMD compensation considerably.

## Appendix

In this section, we will prove a few useful results that are quoted in previous sections of this paper. Most of them are related to the exponential expansion form of the Jones matrix.

For a lossless/gainless linear system, its transmission characteristics can be described by a $2\times 2$ unitary matrix—the Jones matrix [15], which can be expanded in an infinite product of matrix exponentials [16]

$$\mathbf{T}(\omega)=\mathbf{T}_0\exp\left[\Delta\omega\mathbf{M}^{(1)}\right]\exp\left[\frac{1}{2!}\Delta\omega^2\mathbf{M}^{(2)}\right]\times\exp\left[\frac{1}{3!}\Delta\omega^3\mathbf{M}^{(3)}\right]\cdots \tag{44}$$

where $\mathbf{T}_0$ is the Jones matrix at the carrier freqency $\omega_0$, $\Delta\omega:=\omega-\omega_0$ the frequency detuning, and $\mathbf{M}^{(k)}$'s some $2\times 2$ complex matrices to be determined. In the following we shall impose the addition condition that $\det\mathbf{T}=1$. It can be shown that $\mathbf{M}^{(k)}$'s are traceless and skew-Hermitian. It follows that the two eigenvalues of $\mathbf{M}^{(k)}$ are purely imaginary and can be written as

$\pm i\lambda_k$ for some nonnegative $\lambda_k$. In other words, there is some unitary transformation $\mathbf{Q}_k$ such that

$$\mathbf{M}^{(k)} = \mathbf{Q}_k^\dagger \begin{bmatrix} i\lambda_k & 0 \\ 0 & -i\lambda_k \end{bmatrix} \mathbf{Q}_k = \lambda_k \mathbf{Q}_k^\dagger i\sigma_3 \mathbf{Q}_k \tag{45}$$

where $\sigma_3$ is the third of the Pauli matrices, which are defined by

$$\sigma_1 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad \sigma_2 = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}, \quad \sigma_3 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}.$$

Since $i\sigma_3$ is unitary it follows that $\mathbf{M}^{(k)}$ is proportional to a unitary, with proportionality coefficient $\lambda_k$. The first several $\mathbf{M}^{(k)}$s are calculated in [16]:

$$\begin{aligned} \mathbf{M}^{(1)} &= \mathbf{T}_0^\dagger \mathbf{T}' \Big|_{\omega_0} \\ \mathbf{M}^{(2)} &= \mathbf{T}_0^\dagger \mathbf{T}'' \Big|_{\omega_0} - \left(\mathbf{M}^{(1)}\right)^2. \end{aligned} \tag{46}$$

This exponential expansion form (EEF) offers an alternative definition of PMD: the $k$th order PMD is defined by $2\lambda_k$. It also generalizes the concept of principal states of polarization (PSP's) to higher orders: the PSP's of the $k$th order PMD are the eigenvectors of $\mathbf{M}^{(k)}$. In what follows, we demonstrate the relation between the EEF and the conventional formulation of PMD. In particular, we shall establish the identities

$$2\lambda_1 = |\mathbf{\Omega}|, \quad 2\lambda_2 = |\mathbf{\Omega}_\omega| \tag{47}$$

where $\mathbf{\Omega}$ is the PMD vector and $\mathbf{\Omega}_\omega$ its first order derivative with frequency.

Employing the Dirac notation in quantum mechanics, we express the transverse optical field vector as $|\varphi\rangle = (E_x, E_y)^t$ where the superscript "$t$" means matrix transpose. By definition of the Jones matrix

$$|\varphi\rangle = \mathbf{T}(\omega)|\varphi_0\rangle. \tag{48}$$

Fixing the input field $|\varphi_0\rangle$, we have

$$\frac{\partial}{\partial\omega}|\varphi\rangle = \mathbf{T}'(\omega)|\varphi_0\rangle = \mathbf{T}'\mathbf{T}^\dagger|\varphi\rangle, \tag{49}$$

$$\frac{\partial^2}{\partial\omega^2}|\varphi\rangle = \mathbf{T}''(\omega)|\varphi_0\rangle = \mathbf{T}''\mathbf{T}^\dagger|\varphi\rangle. \tag{50}$$

It is easy to show that the matrix $\mathbf{T}'\mathbf{T}^\dagger$ is traceless and skew-Hermitian, therefore there exists a unique real vector $\mathbf{a}$ such that

$$\mathbf{H} := \mathbf{T}'\mathbf{T}^\dagger = -i\mathbf{a}\cdot\boldsymbol{\sigma} \tag{51}$$

where $\boldsymbol{\sigma} = \sigma_1\hat{e}_1 + \sigma_2\hat{e}_2 + \sigma_3\hat{e}_3 = \sigma_i\hat{e}_i$. Here we have also adopted Einstein's summation convention that repeated (dummy) indices are implicitly summed over 1, 2, and 3.

The Poincaré sphere representation of the optical field vector in a fiber, the Stokes vector $\boldsymbol{s}$, is related to $|\varphi\rangle$ by [11], [20]

$$\boldsymbol{s} = \langle\varphi|\boldsymbol{\sigma}|\varphi\rangle. \tag{52}$$

Note that we adopt a slightly different convention here.[2] Taking the frequency derivative of (52) gives

$$\frac{\partial}{\partial\omega}\boldsymbol{s} = \left(\frac{\partial}{\partial\omega}\langle\varphi|\right)\boldsymbol{\sigma}|\varphi\rangle + \langle\varphi|\boldsymbol{\sigma}\left(\frac{\partial}{\partial\omega}|\varphi\rangle\right) = i\langle\varphi|[\boldsymbol{a}\boldsymbol{\sigma}, \boldsymbol{\sigma}]|\varphi\rangle \tag{53}$$

where $[A, B] = AB - BA$ is the commutator of $A$ and $B$. In arriving at (53) we have made use of (49) and (51), and skew-Hermiticity of $\mathbf{H}$. By the well-known commutation relation of Pauli matrices

$$[\sigma_j, \sigma_k] = 2i\epsilon_{jkm}\sigma_m \tag{54}$$

(53) reduces to

$$\frac{\partial\boldsymbol{s}}{\partial\omega} = \langle\varphi|2\boldsymbol{a}\times\boldsymbol{\sigma}]|\varphi\rangle = 2\boldsymbol{a}\times\boldsymbol{s}. \tag{55}$$

Comparing (55) with the definition of PMD vector [21], [22] $(\partial\boldsymbol{s}/\partial\omega) = \mathbf{\Omega}\times\boldsymbol{s}$, we identify $\boldsymbol{a} = \mathbf{\Omega}/2$ and obtain the useful identity

$$\mathbf{T}'\mathbf{T}^\dagger = -\frac{i}{2}\mathbf{\Omega}\boldsymbol{\sigma}. \tag{56}$$

It follows that (49) can be rewritten as

$$\frac{\partial}{\partial\omega}|\varphi\rangle = -\frac{i}{2}\mathbf{\Omega}\boldsymbol{\sigma}|\varphi\rangle. \tag{57}$$

By taking the frequency derivative of the above equation we obtain

$$\frac{\partial^2}{\partial\omega^2}|\varphi\rangle = -\frac{i}{2}\mathbf{\Omega}_\omega\boldsymbol{\sigma}|\varphi\rangle + \left(-\frac{i}{2}\mathbf{\Omega}\boldsymbol{\sigma}\right)^2|\varphi\rangle \tag{58}$$

Comparing (49) with (57), (50) with (58) at $\omega = \omega_0$, and making use of (46), we identify:

$$\mathbf{T}_0\mathbf{M}^{(1)}\mathbf{T}_0^\dagger = -\frac{i}{2}\mathbf{\Omega}\boldsymbol{\sigma}\Big|_{\omega_0} \tag{59}$$

$$\mathbf{T}_0\mathbf{M}^{(2)}\mathbf{T}_0^\dagger = -\frac{i}{2}\mathbf{\Omega}_\omega\boldsymbol{\sigma}\Big|_{\omega_0}. \tag{60}$$

Taking the determinant on both sides of (59) and (60), and using

$$\det|\boldsymbol{r}\boldsymbol{\sigma}| = -\boldsymbol{r}^2, \quad \det\left|\mathbf{M}^{(k)}\right| = \lambda_k^2 \tag{61}$$

we obtain

$$2\lambda_1 = |\mathbf{\Omega}|_{\omega=\omega_0}, \quad 2\lambda_2 = |\mathbf{\Omega}_\omega|_{\omega=\omega_0}. \tag{62}$$

Since $\mathbf{\Omega}$ and $\mathbf{\Omega}_\omega$ are random quantities whose time average is equivalent to the frequency average [11], we can relax the con-

[2]The conventionally defined Stokes vector, $\boldsymbol{s}^0$, differs from our definition by a cyclic permutation of basis: $\boldsymbol{s}_1 = \boldsymbol{s}_2^0$, $\boldsymbol{s}_2 = \boldsymbol{s}_3^0$, $\boldsymbol{s}_3 = \boldsymbol{s}_1^0$. Consequently the PMD vector in the following analysis also differs from the conventionally defined PMD vector $\mathbf{\Omega}^0$ by a change of basis: $\mathbf{\Omega}_1 = \mathbf{\Omega}_2^0$, $\mathbf{\Omega}_2 = \mathbf{\Omega}_3^0$, $\mathbf{\Omega}_3 = \mathbf{\Omega}_1^0$. This transformation of basis obvious produces no physical difference in the results.

dition $\omega = \omega_0$ in (62) when dealing with their average values. In this sense, we write

$$2\lambda_1 = |\mathbf{\Omega}|, \quad 2\lambda_2 = |\mathbf{\Omega}_\omega|\,. \tag{63}$$

These identities are used repeatedly in the main text of this paper.

Finally, we note that the useful identity (56) also offers a simple proof of the following relation:

$$\tfrac{1}{4}\,\Omega^2 = \det|\mathbf{T}'|. \tag{64}$$

By taking the determinant on both sides of (56), (64) is a direct corollary of the first identity in (61). This result is used in (29) of the main text.

## ACKNOWLEDGMENT

The authors thank the first reviewer for clarifying a few concepts in an earlier version of this paper.

## REFERENCES

[1] G. J. Foschini and C. D. Poole, "Statistical theory of polarization dispersion in single mode fibers," *J. Lightwave Technol.*, vol. 9, pp. 1439–1456, Nov. 1991.
[2] N. Gisin, "Solutions of the dynamical equation for polarization dispersion," *Opt. Commun.*, vol. 86, pp. 371–373, 1991.
[3] P. K. A. Wai and C. R. Menyuk, "Polarization decorrelation in optical fibers with randomly varying birefringence," *Opt. Lett.*, vol. 19, pp. 1517–1519, 1994.
[4] ——, "Polarization mode dispersion, decorrelation,and diffusion in optical fibers with randomly varying birefringence," *J. Lightwave Technol.*, vol. 14, pp. 148–157, Feb. 1996.
[5] P. Ciprut, B. Gisin, N. Gisin, R. Passy, J. P. Von der Weid, F. Prieto, and C. W. Zimmer, "Second order polarization mode dispersion: Impact on analog and digital transmissions," *J. Lightwave Technol.*, vol. 16, pp. 757–771, May 1998.
[6] L. Arnold, *Stochastic Differential Equations: Theory and Applications*. New York: Wiley, 1974.
[7] W. Horsthemke and R. Lefever, *Noise Induced Transitions*. Berlin, Germany: Springer, 1984.
[8] B. Øksendal, *Stochastic Differential Equations: An Introduction with Applications*. New York: Springer, 1998.
[9] H. J. Carmichael, *Statistical Methods in Quantum Optics*. New York: Springer, 1999.
[10] Y. Li and A. Yariv, "Solutions of the dynamical equation for polarization-mode dispersion in presence of polarization-dependent losses," J. Opt. Soc. Amer. B., submitted for publication.
[11] N. Gisin and J. P. Pellaux, "Polarization mode dispersion: Time versis frequency domains," *Opt. Commun.*, vol. 89, pp. 316–323, 1992.
[12] F. Corsi, A. Galtarossa, and L. Palmieri, "Polarization mode dispersion characterization of single-mode optical fiber using backscattering technique," *J. Lightwave Technol.*, vol. 16, pp. 1832–1843, Oct. 1998.
[13] M. Karlsson and J. Brentel, "Autocorrelation function of the polarization-mode dispersion vector," *Opt. Lett.*, vol. 24, no. 14, pp. 939–941, 1999.
[14] C. D. Poole and D. L. Favin, "Polarization-mode dispersion measurements based on transmission spectra through a polarizer," *J. Lightwave Technol.*, vol. 12, pp. 917–929, June 1994.
[15] R. C. Jones, "A new calculus for the treatment of optical systems (i)," *J. Opt. Soc. Amer.*, vol. 31, pp. 488–493, 1941.
[16] A. Eyal, W. K. Marshall, M. Tur, and A. Yariv, "A new representation of second order polarization mode dispersion," *Electron. Lett.*, vol. 35, no. 19, pp. 1658–1659, 1999.
[17] R. Noé, D. Sandel, M. Yoshida-Dierolf, S. Hinz, C. Glingener, C. Scheerer, A. Schöpflin, and G. Fischer, "Polarization mode dispersion compensation at 20 gbit/s with fiber-based distributed equaliser," *Electron. Lett.*, vol. 34, no. 25, pp. 2421–2422, 1998.
[18] R. Noé, D. Sandel, M. Yoshida-Dierolf, S. Hinz, V. Mirvoda, A. Schöpflin, C. Glingener, E. Gottwald, C. Scheerer, G. Fischer, T. Weyrauch, and W. Haase, "Polarization mode dispersion compensation at 10, 20, and 40 gb/s with various optical equalizers," *J. Lightwave Technol.*, vol. 17, pp. 1602–1616, Sept. 1999.
[19] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York, NY: Wiley, 1991.
[20] N. J. Frigo, "A generalized geometrical representation of coupled mode theory," *IEEE J. Quantum Electron.*, vol. 22, no. 11, pp. 2131–2140, 1986.
[21] C. D. Poole, N. S. Bergano, R. E. Wagner, and H. J. Schulte, "Polarization dispersion and principal states in a 147 km undersea lightwave cable," *J. Lightwave Technol.*, vol. 6, pp. 1185–1190, July 1988.
[22] D. Andresciani, F. Curti, F. Matera, and B. Daino, "Measurement of the group-delay difference between the principal states of polarization on a low-birefringence terrestrial fiber cable," *Opt. Lett.*, vol. 12, no. 10, pp. 844–846, 1987.

**Yi Li**, photograph and biography not available at the time of publication.

**Avishay Eyal** (S'96–A'98), photograph and biography not available at the time of publication.

**Amnon Yariv** (S'56–M'59–F'70–LF'95), photograph and biography not available at the time of publication.